



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/814,081	03/31/2004	Wu Chou	633-001us	5517
47912	7550	02/02/2011		
Avaya			EXAMINER	
DEMONT & BREYER, LLC			YEN, ERIC L.	
100 COMMONS WAY, STE 250				
HOLMDEL, NJ 07733				
			ART UNIT	PAPER NUMBER
			2626	
			NOTIFICATION DATE	DELIVERY MODE
			02/02/2011	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docketing@dblaw.com

Office Action Summary

Application No.

10/814,081

Applicant(s)

CHOU ET AL.

Examiner

ERIC YEN

Art Unit

2626

Period for Reply -- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 21 January 2011.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1, 3, 5-7 and 9-18 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1, 3, 5-7, 9-18 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
- Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- 1) ☐ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-940)
- 3) ☐ Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date _____
- 4) ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date _____
- 5) ☐ Notice of Informal Patent Application
- 6) ☐ Other: _____

DETAILED ACTION

Response to Amendment

1. In response to the Office Action mailed 11/18/10, applicant has submitted an amendment filed 1/21/11.

Claims 1, 3, 7, 19, 12, 14, 15, 16, 18, have been amended. Claim 4 has been cancelled.

Response to Arguments

1. Applicant's arguments filed 1/21/11 have been fully considered but they are not persuasive.

Applicant argues that "the use of a combination of terms from multiple and diverse kinds of sources, and particularly the combination of word-terms and word-classes, provides the present invention with added robustness and significant performance improvements over the prior art" and argues that there is no motivation to combine the references (Amendment, page 13).

Applicant's argument, however, does not apply in the current rejection where motivation to combine is not required. The factual inquiries involving simple substitution of one element with another only requires what **one of ordinary skill in the art COULD have done** (see factual elements of obviousness rationales in MPEP 2143). What one COULD do does not involve doing something for any particular advantage and more particularly does not need to be APPLICANT'S advantage (e.g. one could mix chemicals that can result in an explosion but not necessarily would for potential harm).

KSR describes that the prior art need not be directed to the same problem as applicant's. As long as the claimed limitations COULD be combined to obtain predictable results, the prior art combination is proper.

Applicant then argues that "template in hindsight" strategy is impermissible when applying a combination where each reference contributes a portion of information to the combination".

2. In response to applicant's argument that the examiner's conclusion of obviousness is based upon improper hindsight reasoning, it must be recognized that any judgment on obviousness is in a sense necessarily a reconstruction based upon hindsight reasoning. But so long as it takes into account only knowledge which was within the level of ordinary skill at the time the claimed invention was made, and does not include knowledge gleaned only from the applicant's disclosure, such a reconstruction is proper. See *In re McLaughlin*, 443 F.2d 1392, 170 USPQ 209 (CCPA 1971).

The previous paragraph (form paragraph) aside, hindsight analysis applies to the MOTIVATION to combine being identical to applicant's reason for combining the elements. It does not apply to using the claim to find the references. The substance of the prior art is what it is. If it falls within a broad, reasonable interpretation the claim limitations then it meets the claim limitations. It also need not meet individually every claim limitation (i.e. need not anticipate the claims individually which as a practical matter would not make combination necessary). The mere fact that the prior art was found using applicant's claim language does not mean that impermissible hindsight is

applied in the combination, when the factual inquiries describing whether they could be combined are still met.

Applicant argues that "whereas the applicants have stressed that a significant advantage to the present invention is derived from using the combination of terms that comprises both word-terms and class-terms.... The Office has failed to cite a reference that teaches, suggests, or motivates such a combination" and "instead, the Office uses the pending claim as a template and the Office combines two disparate references to create the very element that confers a major advantage to the present invention" (Amendment, page 14).

This argument, however, is irrelevant because KSR explicitly described that the "teaching, suggestion, and motivation" test was not the only test for combination. No teaching, suggestion, or motivation was applied in the rejection, only that one of ordinary skill in the art could have substituted the data sources which Li-2002 draws the words and categories/classes from in order to obtain predictable results. Since Li-2002 makes selections from a data source that includes words and word-classes and Diab teaches a corpus including the same type of information (words and word-classes/word-senses), there is something in common with whatever Li selects words/word-classes from and the corpus in Diab. Specifically, they both teach data sources incorporating words and word-classes, and thus are not as disparate as applicant alleges they are. The result of substituting the Diab corpus with Li's word/word-class source (whatever it may be) is predictable because Li explicitly teaches the result. Particularly, Li teaches a system that selects word/word-class information from a source of information includes

that information. Things do not get much more predictable than when they are spelled out in the references.

Applicant's next argument is that "selecting" is inextricably tied into "combining... that results in a combination of terms" arguing that "the availability of both word-terms and class-terms as possible candidates for selection is a meaning that neither Diab nor Li-2002 contemplate (Amendment, page 14).

Applicant, however, appears to have confused "availability" of selection with "actual" selection, and as claimed, there is no requirement that "plurality of terms" needs to include both word terms and word-classes. As claimed only the combination of terms needs to include word-terms and word-classes. Just as there is no implicit requirement that word terms and word-classes must be derived from separate and distinct sources (i.e. not a corpus which ALREADY has both pieces of information), there is no requirement that "plurality of terms" necessarily include both word-terms and word-classes merely because they are derived from the combination of terms. "Plurality of terms" "wherein a term is one of a word-term and a word-class" only requires that there be, plainly based on the claim language, multiple word-terms OR word-classes (necessarily OR because "one of a word-term and a word-class" only requires one or the other). "Plurality" and "combination" may both share the characteristic of having more than one data entity, but they are, as claimed, separate and distinct entities with different limits on interpretation (i.e. only "combination" as defined in the claims requires both word-terms and word-classes, while "plurality" just has to be one or the other). Plurality does not inherit the characteristics of combination just because they are both

"of terms" when "terms" are only ONE OF "word terms" and "word classes". Therefore, as applied in the rejection below, because at the very least multiple WORDS are selected (where the words selected in Li-2002 are multiple/plurality of words/"terms that are one of a word-term and a word class"), the claim limitation is met. Merely claiming the word "and" does not mean that BOTH of the claimed entities before and after and must be part of anything involving the claim language "term" especially when "term" is defined as only ONE OF those things.

As far as "availability" for selection, Diab's corpus has word-class information (word senses) assigned to their corresponding words (word terms). Something is "available" merely because it exists. It is also obvious that the data is "available for selection" because if nobody can do anything with the word-classes, then the information serves no purpose and Diab describes word-classes as being directed to a particular purpose, or at least one of ordinary skill in the art would recognize that word senses as commonly used in the art are used for disambiguation (e.g. "server" in a machine sense or a restaurant sense). Therefore, it is also not true that "availability" (i.e. something existing so that it may be selected) is not taught in the prior art. This is distinct from ACTUALLY selecting word-terms and word-classes.

Applicant then argues that it would be "predictable to one skilled in the art that the method of the pending claim achieves advantageous results" because "it is speculative and conclusory" and "combining two different kinds of terms (word-terms and word-classes) is not taught in the cited references and there is no reason to predict that the results would achieve the performance improvements cited by the present

inventors" and "the results could just as easily have been the same or worse, or costlier, or otherwise disadvantageous" and "it is just as conceivable that a combination of terms would result in a more 'confused' outcome, because of the diverse kinds of terms being combined together (Amendment, page 15).

Predictable results, however, does not require any reason or motivation to combine nor does it, as discussed above, have to be significantly improved. All that is required is that the results are predictable, and as discussed above, it is hard for anything to be more predictable than when they are taught by the references. Li teaches obtaining information from a word/word-class data entity, and Diab teaches a word/word-class/sense data entity. Also, Diab's corpus already combines word-terms and word-classes to produce a combination of the two (sense-tagged word corpus), and thus the information is not so disparate that it would result in a "confusing" result (i.e. it cannot fall under "confusing" if it's already taught as being combined in the prior art because somebody was obviously not confused and knew what they were doing when they made the combination of terms/corpus).

Applicant argues that "moreover, asserting that because a reference cites 'selection from something,' 'some sort of analysis... is possible' hardly provides a reasoned and clear articulation of why the claimed limitations lack inventiveness" (Amendment, page 15).

The mere use of general terms, however, does not render an articulation unclear, especially when they indicate a commonality between substituted components that is not just "it exists". As described above in the arguments and below in the rejection, Li's

selection source and Diab both contain "word classes" and "word terms". Li's is (Section 3, paragraph 2) a training corpus with categories labeled, and Diab's is a corpus with sense labels (category labels). This is made clear in the articulation of the rationale which states/stated, in relevant part **"the substitution of** a corpus containing word class information and word term information generated by some means used by Li to generated a matrix, with another corpus containing the same information derived/generated by a processor. Diab teaches that a corpus generated by a processor and containing word class information and word term information **was known in the art**". Therefore, it was previously clearly articulated that it was the corpus being substituted with another corpus containing word and word class/word sense information, and applicant's assertion on page 15 of the Remarks is incorrect.

Finally, applicant argues that "determining from the matrix, based on joint classification of the word" and "generated from an automatic word-class clustering algorithm" (previous claim 4) is not taught in the prior art (Amendment, page 15).

Whatever applicant meant by "joint classification", however, is not what the scope is limited to, just as "word-class" is not defined with any particular specificity in the Specification or in the claims, nor do either of these terms incorporate any common meaning that limits it to any particular subset of part-of-speech labels/word senses/topics/contexts or any other subset of everything known in the art. "Joint classification" is, as previously described and repeated below in the rejection of claim 3 "is "joint classification" in the sense that it "jointly" uses both term information and category information to perform classification [i.e. "joint classifier is configured to

determine at least one category for the words, by applying a combination of word information and word class information to the words”, Specification, page 6]”.

Therefore, under the same rationale, and unless applicant further defines joint classification to be exactly what applicant intends it to be in the claims, the rejection is proper.

Also, applicant has not addressed the automatic word clustering taught in Sakai, which provides an alternative method of generating categories/classes (i.e. by automatic word clustering). Diab's and Li's corpora are generated somehow because data does not appear out of thin air. Since the end result shares similar characteristics (text/words assigned categories/classes), whatever method was used to generate the Diab/Li corpora can be substituted with an automatic word clustering to produce the same result (a data entity which includes word sense and word class information). Again, the mere use of “whatever” does not render the articulation unclear because the common aspect of what is taught in the different references is established.

Therefore, the examiner maintains similar prior art rejections to those previously presented, adjusted for the amendments to the claims.

Claim Rejections - 35 USC § 103

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the

invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

2. Claims 1, 3, 5-7, 9-18, are rejected under 35 U.S.C. 103(a) as being unpatentable over Li et al. ("Improving Latent Semantic indexing Based Classifier With Information Gain"), hereafter Li, in view of Diab et al. ("An Unsupervised Method for Word Sense Tagging using Parallel Corpora"), hereafter Diab.

As per Claim 1, 15, Li teaches a method (and corresponding apparatus, where the joint classifier is defined in the claim by its function identical to the method in Claim 1) comprising: identifying, by a processor-based device, a word that (i) is received in a communication, and (ii) is natural-language word, wherein the processor-based device is to determine a category for the communication ("LSI classifier... categorize an unknown document... derived from... as in LSI according to IG enhanced term-document matrix... similarity... n-best categories", Section 3; "natural language understanding... directing the user's call... matches a user's request", Section 4, Experimental Setup, paragraph 1; where users communicate by speaking in natural language which includes speaking natural language words; "natural language understanding... directing the user's call... matches a user's request", Section 4, Experimental Setup, paragraph 1; where to match a request when a request is natural language speech, it is at least obvious that words representing semantic information used to service the request are identified)

selecting by the processor-based device a plurality of terms wherein the selecting is based on an information-gain value those terms that correspond to the word ("term-

document matrix... each selected term... IG based term selection is implemented... terms are selected and used in the term-document matrix based on their discriminative power", Section 3; where the terms selected are IG based and sorted by their individual values. Applicant does not claim that only those terms that correspond to the word are selected, and so selecting a larger set of terms which happens to include terms that correspond to the word also reads on the claim language, and it is obvious that these terms are selected because Li teaches categorizing the word is performed, which cannot be done if the terms corresponding to the word were not selected)

generating by the processor-based device a matrix, wherein (i) the matrix comprises a plurality of categories and a plurality of terms, and (ii) each term in the matrix is associated with at least one category ("term-document matrix M is formed by terms... selected term is mapped to a unique row vector and each category is mapped to a unique column vector", Section 3, especially paragraph 3; where the matrix is formed by combining terms [i.e. word terms] and categories [word-classes], and in a matrix, the matrix cell corresponding to a specific row's term/word-term and a specific column's category/word-class associates the term and the category corresponding to a cell; Alternatively, words are naturally associated with some particular class [e.g. words that are verbs, medical words, English words, etc.] and "at least one category" as claimed does not necessarily refer to any category in particular so any word naturally reads on this claim limitation because words are, by virtue of what they are, part of some form of category)

determining from the matrix, based on a joint classification of the word by the processor-based device, a category for the word ("LSI classifier... categorize an unknown document... derived from... as in LSI according to IG enhanced term-document matrix... similarity... n-best categories", Section 3; "user's request", Section 4; where categorizing a document by consequence categorizes the word in that document, and this is "joint classification" in the sense that it "jointly" uses both term information and category information to perform classification [i.e. "joint classifier is configured to determine at least one category for the words, by applying a combination of word information and word class information to the words", Specification, page 6])).

Li fails to teach combining: (i) at least one set of word-terms and (ii) at least one set of word-classes, wherein a term is one of a word-term and a word-class, where the plurality of terms selected is from the combination of terms, where those terms are in the combination of terms, and wherein the combining results in a combination of terms

Diab teaches/suggests combining: (i) at least one set of word-terms and (ii) at least one set of word-classes, wherein a term is one of a word-term and a word-class, ("word sense tagging... automatically sense annotating... large amounts of data... using an unsupervised algorithm... bootstrap... creating a sense-tagged corpus", Introduction, especially paragraph 3; "project the sense tags from the target side to the source side... KIND-OF-DRAMA sense... CALAMITY... the tagging... would yield... large number of French words will receive tags from the English sense inventory", Approach, especially 4th bullet, paragraph ending at the upper right of page 257, and last paragraph; Applicant only claims identification but does not specify where the identification limits

anything else in the claims, so as long as the generated data includes the word that also exists in the communication, it is "based on the word" that the communication "comprises"; Diab teaches sense-tagging words to create a corpus of sense-tagged data. Each of these sense-tagged words includes a word-class like CALAMITY and a word-term like catastrophe. Diab teaches the existence of different senses [like KIND-OF-DRAMA] and it is at least obvious that catastrophe is not the only word with different senses in French. The sense-annotated words in the corpus, collectively, are a "combination of terms" because the classes/senses are combined with their corresponding words, and at the very least in set theory sets can include 1 element, or alternatively, collectively all of the sense-annotated words are a set of word-terms and their corresponding senses collectively are a set of word-classes)

where the plurality of terms selected is from the combination of terms, where those terms are in the combination of terms, and wherein the combining results in a combination of terms ("word sense tagging... automatically sense annotating... large amounts of data... using an unsupervised algorithm... bootstrap... creating a sense-tagged corpus", Introduction, especially paragraph 3; "project the sense tags from the target side to the source side... KIND-OF-DRAMA sense... CALAMITY... the tagging... would yield... large number of French words will receive tags from the English sense inventory", Approach, especially 4th bullet, paragraph ending at the upper right of page 257, and last paragraph; where Li [in Section 3, especially paragraph 2] teaches that the term-document matrix is generated from a labeled corpus, though it does not specifically state how the corpus is generated. Diab teaches generating a labeled

corpus by a processor which contains the very information that Li wishes to extract for Li's matrix. Therefore, one of ordinary skill in the art can simply substitute the corpus used in Li with one generated from a processor as per Diab that contains the information needed to generate the matrix, and which can be used to bootstrap the classifier described in Li)

Therefore, it would have been obvious to one of ordinary skill in the art at the time of invention to perform a **simple substitution** of one corpus which the information-gain-matrix performs selection from with another, because Li teaches a classification method/device **which differed from the claimed device by the substitution of a** corpus containing word class information and word term information generated by some means used by Li to generate a matrix, with another corpus containing the same information derived/generated by a processor. Diab teaches that a corpus generated by a processor and containing word class information and word term information **was known in the art. One of ordinary skill in the art could have substituted one** corpus **for another to obtain the predictable results of** a system which performs classification using a matrix generated from a corpus containing terms and classes (as per Li) where the corpus containing terms and classes is generated by a processor (as per Diab).

Li, in view of Diab, fail to teach wherein an automatic word class clustering algorithm is utilized to generate the word-classes.

Sakai teaches wherein an automatic word class clustering algorithm is utilized to generate the word-classes ("category decision rules... each text is classified to a

category according to the category decision rule", col. 3, lines 35-50; "automatically creates a new category", col. 6, line 53 – col. 7, line 5; "if a cluster consisting of a large number of texts... new category to which this cluster is classified", col. 6, lines 34-40; "cluster generation unit", col. 6, lines 7-24; where the clustering is automatically performed and whose results is used for a new word class, and so it is an automatic word class clustering algorithm and is used to generate new word class [i.e., category] rules/information. Li and Diab teach where categories/senses are generated somehow [since they must have been derived from somewhere to be used in tagging], without providing the specifics. Sakai teaches another method for generating the same data and so a simple substitution of the generation can be performed to yield the class/sense information used in Diab's sense tagging).

Therefore, it would have been obvious to one of ordinary skill in the art to perform a **simple substitution** of one category with another, because Li and Diab teach a **device which differed from the claimed device by the substitution of categories** generated by some unspecified manner with categories generated by clustering. Sakai teaches that categories generated by clustering **were known in the art. One of ordinary skill in the art could have substituted one known element for another** by using categories generated from clustering instead in the unsupervised tagging in Diab **in order to obtain the predictable results of a system that performs classification** based on a matrix derived from word class and word term data (Li) where the word class and word term data are automatically generated by a processor (Diab) based on classes determined initially using a form of clustering (Sakai).

As per Claim 3, 16, 18, Li teaches/suggests (along with its apparatus equivalent of Claim 16 and article of manufacture equivalent in Claim 18, where claim 18 includes the limitations of both Claims 1 and 3 and so incorporates the rejections presented above regarding claim 1 as well) routing the communication by a communication system to a particular one of a plurality of destination terminals of the communication system, wherein the routing is based on the category, and wherein the communication system comprises the processor-based device and the plurality of destination terminals ("routing... appropriate destination within a call center", Section 4; where the routing to destinations in the call-center/system based on the query's categorization, which is done via the joint classifier, which includes categorizing the words in the query, and the "system" can be interpreted as the call router and all of the places that the call is routed to; and this is "joint classification" in the sense that it "jointly" uses both term information and category information to perform classification [i.e. "joint classifier is configured to determine at least one category for the words, by applying a combination of word information and word class information to the words", Specification, page 6]).

As per Claim 5, Li teaches selecting of the plurality of terms is further based on a percentile value applied to the respective information-gain values of each term in the combination of terms ("top p percentile... according to the IG score", Section 3; where the terms being in the combination of terms is addressed in the same manner as above

by Diab in the parent claim, the set selected which is part of the corpus they are selected from can be interpreted as the combination of terms as well).

As per Claim 6, Li teaches wherein the information-gain value for each term in the combination of terms, indicates the average entropy variations over a plurality of possible categories for each term in the combination of terms ("significance of the term based on the entropy variations of the categories, which relates to the perplexity of the classification task", Section 2; "literal terms... may not match those of a relevant document", Section 1, paragraph 1; "IG enhanced... classified... categorize an unknown document", Section 3; where the entropy variations are taught by Li to relate to perplexity and so an entropy calculation is also a perplexity calculation and Equation 1 describes the information gain value being calculated from entropy/perplexity. Also the subscript t_i at the end of Section 2 at least suggests that there is more than one term for which the information gain is calculated; where the terms being in the combination of terms is addressed above by Diab in the parent claim, the set selected which is part of the corpus they are selected from can be interpreted as the combination of terms as well).

As per Claim 7, 17, Li teaches (along with its apparatus equivalent of Claim 17) wherein the category of the word is a cell in the matrix ("cell... j-th category", Section 3).

As per Claim 9, Li fails to teach wherein the combination of terms is generated by interleaving individual word-terms with their corresponding word-classes.

Diab teaches/suggests wherein the combination of terms is generated by interleaving individual word-terms with their corresponding word-classes ("word sense tagging... automatically sense annotating... large amounts of data... using an unsupervised algorithm... bootstrap... creating a sense-tagged corpus", Introduction, especially paragraph 3; "project the sense tags from the target side to the source side... KIND-OF-DRAMA sense... CALAMITY... the tagging... would yield... large number of French words will receive tags from the English sense inventory", Approach, especially 4th bullet, paragraph ending at the upper right of page 257, and last paragraph; where word-sense tagging interleaves [mixes or inserts the sense tags regularly between words in the corpus] the sense/word-classes and their respective words/word-terms)

Therefore, it would have been obvious to one of ordinary skill in the art at the time of invention to perform a **simple substitution** of one corpus which the information-gain-matrix performs selection from with another, because Li teaches a classification method/device **which differed from the claimed device by the substitution of a** corpus containing word class information and word term information generated by some means used by Li to generated a matrix, with another corpus containing the same information derived/generated by a processor. Diab teaches that a corpus generated by a processor and containing word class information and word term information **was known in the art. One of ordinary skill in the art could have substituted one corpus for another to obtain the predictable results of** a system which performs

classification using a matrix generated from a corpus containing terms and classes (as per Li) where the corpus containing terms and classes is generated by a processor (as per Diab).

As per Claim 10, Li teaches/suggests a method comprising: identifying, by a processor-based device, a word that (i) is received in a communication, and (ii) is natural-language word, wherein the processor-based device is to determine a category for the communication ("LSI classifier... categorize an unknown document... derived from... as in LSI according to IG enhanced term-document matrix... similarity... n-best categories", Section 3; "natural language understanding... directing the user's call... matches a user's request", Section 4, Experimental Setup, paragraph 1; where users communicate by speaking in natural language which includes speaking natural language words; "natural language understanding... directing the user's call... matches a user's request", Section 4, Experimental Setup, paragraph 1; where to match a request when a request is natural language speech, it is at least obvious that words representing semantic information used to service the request are identified)

selecting by the processor-based device a plurality of terms wherein the selecting is based on an information-gain value those terms that correspond to the word ("term-document matrix... each selected term... IG based term selection is implemented... terms are selected and used in the term-document matrix based on their discriminative power", Section 3; where the terms selected are IG based and sorted by their individual values)

generating by the processor-based device a term-category matrix, that comprises a plurality of terms and a plurality of categories, wherein each term in the matrix is associated with at least one category ("term-document matrix M is formed by terms... selected term is mapped to a unique row vector and each category is mapped to a unique column vector", Section 3, especially paragraph 3; where the matrix is formed by combining terms [i.e. word terms] and categories [word-classes], and in a matrix, the matrix cell corresponding to a specific row's term/word-term and a specific column's category/word-class associates the term and the category corresponding to a cell; Alternatively, words are naturally associated with some particular class [e.g. words that are verbs, medical words, English words, etc.] and "at least one category" as claimed does not necessarily refer to any category in particular so any word naturally reads on this claim limitation because words are, by virtue of what they are, part of some form of category)

classifying the communication by utilizing a joint classifier upon the at least one word, wherein the joint classifier comprises the term-category matrix ("LSI classifier... categorize an unknown document... derived from... as in LSI according to IG enhanced term-document matrix... similarity... n-best categories", Section 3; "user's request", Section 4; where categorizing a document by consequence categorizes the word in that document, and this is "joint classification" in the sense that it "jointly" uses both term information and category information to perform classification [i.e. "joint classifier is configured to determine at least one category for the words, by applying a combination of word information and word class information to the words", Specification, page 6]).

Li fails to teach combining: (i) at least one set of word-terms and (ii) at least one set of word-classes, wherein a term is one of a word-term and a word-class, where the plurality of terms selected is from the combination of terms, where those terms are in the combination of terms, and wherein the combining results in a combination of terms

Diab teaches/suggests combining: (i) at least one set of word-terms and (ii) at least one set of word-classes, wherein a term is one of a word-term and a word-class ("word sense tagging... automatically sense annotating... large amounts of data... using an unsupervised algorithm... bootstrap... creating a sense-tagged corpus", Introduction, especially paragraph 3; "project the sense tags from the target side to the source side... KIND-OF-DRAMA sense... CALAMITY... the tagging... would yield... large number of French words will receive tags from the English sense inventory", Approach, especially 4th bullet, paragraph ending at the upper right of page 257, and last paragraph; Applicant does not claim that the word that the generation is based on was derived from the communication, so as long as the generated data includes the word that also exists in the communication, it is "based on the word" that the communication "comprises"; Diab teaches sense-tagging words to create a corpus of sense-tagged data. Each of these sense-tagged words includes a word-class like CALAMITY and a word-term like catastrophe. Diab teaches the existence of different senses [like KIND-OF-DRAMA] and it is at least obvious that catastrophe is not the only word with different senses in French. The sense-annotated words in the corpus, collectively, are a "combination of terms" because the classes/senses are combined with their corresponding words, and at the very least in set theory sets can include 1 element, or alternatively, collectively all

of the sense-annotated words are a set of word-terms and their corresponding senses collectively are a set of word-classes)

where the plurality of terms selected is from the combination of terms, where those terms are in the combination of terms, and wherein the combining results in a combination of terms ("word sense tagging... automatically sense annotating... large amounts of data... using an unsupervised algorithm... bootstrap... creating a sense-tagged corpus", Introduction, especially paragraph 3; "project the sense tags from the target side to the source side... KIND-OF-DRAMA sense... CALAMITY... the tagging... would yield... large number of French words will receive tags from the English sense inventory", Approach, especially 4th bullet, paragraph ending at the upper right of page 257, and last paragraph; where Li [in Section 3, especially paragraph 2] teaches that the term-document matrix is generated from a labeled corpus, though it does not specifically state how the corpus is generated. Diab teaches generating a labeled corpus by a processor which contains the very information that Li wishes to extract for Li's matrix. Therefore, one of ordinary skill in the art can simply substitute the corpus used in Li with one generated from a processor as per Diab that contains the information needed to generate the matrix, and which can be used to bootstrap the classifier described in Li)

Therefore, it would have been obvious to one of ordinary skill in the art at the time of invention to perform a **simple substitution** of one corpus which the information-gain-matrix performs selection from with another, because Li teaches a classification method/device **which differed from the claimed device by the substitution of a**

corpus containing word class information and word term information generated by some means used by Li to generated a matrix, with another corpus containing the same information derived/generated by a processor. Diab teaches that a corpus generated by a processor and containing word class information and word term information **was known in the art. One of ordinary skill in the art could have substituted one corpus for another to obtain the predictable results** of a system which performs classification using a matrix generated from a corpus containing terms and classes (as per Li) where the corpus containing terms and classes is generated by a processor (as per Diab).

Li, in view of Diab, fail to teach wherein an automatic word class clustering algorithm is utilized to generate the word-classes.

Sakai teaches wherein an automatic word class clustering algorithm is utilized to generate the word-classes ("category decision rules... each text is classified to a category according to the category decision rule", col. 3, lines 35-50; "automatically creates a new category", col. 6, line 53 – col. 7, line 5; "if a cluster consisting of a large number of texts... new category to which this cluster is classified", col. 6, lines 34-40; "cluster generation unit", col. 6, lines 7-24; where the clustering is automatically performed and whose results is used for a new word class, and so it is an automatic word class clustering algorithm and is used to generate new word class [i.e., category] rules/information. Li and Diab teach where categories/senses are generated somehow [since they must have been derived from somewhere to be used in tagging], without providing the specifics. Sakai teaches another method for generating the same data

and so a simple substitution of the generation can be performed to yield the class/sense information used in Diab's sense tagging).

Therefore, it would have been obvious to one of ordinary skill in the art to perform a **simple substitution** of one category with another, because Li and Diab teach a **device which differed from the claimed device by the substitution of** categories generated by some unspecified manner with categories generated by clustering. Sakai teaches that categories generated by clustering **were known in the art. One of ordinary skill in the art could have substituted one known element for another** by using categories generated from clustering instead in the unsupervised tagging in Diab **in order to obtain the predictable results of** a system that performs classification based on a matrix derived from word class and word term data (Li) where the word class and word term data are automatically generated by a processor (Diab) based on classes determined initially using a form of clustering (Sakai).

As per Claim 11, Li teaches wherein a cell I,j , of the term-category matrix represents a classification by the processor-based device of an i -th selected term into a j -th category ("LSI classifier... categorize an unknown document... similarity... n -best categories", Section 3; "user's request", Section 4; where categories in the matrix are among the j categories and categorizing a request including terms categorizes it into a category among the categories numbered by the values of j . Cells of matrices correspond to a particular row and column and in the case of a cell corresponding to a

word/row and class/column, the cell represents an association between a word and the corresponding class)

As per Claim 12, Li teaches/suggests a method comprising: identifying, by a processor-based device, a word that (i) is received in a communication, and (ii) is natural-language word, wherein the processor-based device is to determine a category for the communication ("LSI classifier... categorize an unknown document... derived from... as in LSI according to IG enhanced term-document matrix... similarity... n-best categories", Section 3; "natural language understanding... directing the user's call... matches a user's request", Section 4, Experimental Setup, paragraph 1; where users communicate by speaking in natural language which includes speaking natural language words; "natural language understanding... directing the user's call... matches a user's request", Section 4, Experimental Setup, paragraph 1; where to match a request when a request is natural language speech, it is at least obvious that words representing semantic information used to service the request are identified)

selecting by the processor-based device a plurality of terms wherein the selecting is based on an information-gain value those terms that correspond to the word ("term-document matrix... each selected term... IG based term selection is implemented... terms are selected and used in the term-document matrix based on their discriminative power", Section 3; where the terms selected are IG based and sorted by their individual values)

wherein the selecting comprises: i) calculating an information gain value for each term that corresponds to the word ("terms are selected and used... according to IG criterion... sort the terms", Section 3; "terms in documents", Section 1, paragraph 1; where sorting terms by their IG values means that each term had its IG value calculated such that they can be sorted, and the terms are in documents that communicate information [received at the input to the classification system], and documents contain words and so the terms in this context are words)

ii) sorting the terms in the union of terms in a descending order of information gain values ("sort the terms by their IG values in descending order", Section 3)

iii) setting a threshold of an information gain value corresponding to a specified percentile ("select top p percentile of terms according to the IG score distribution", Section 3; where taking the top p percentile sets the lowest of that p percentile as the threshold IG score)

iv) selecting only the terms having an information gain value greater than or equal to the threshold to generate the plurality of terms ("select top p percentile of terms", Section 3; where taking the top p percentile takes all terms exceeding the lowest IG value in that percentile and excludes everything falling below the percentile).

Li fails to teach combining: (i) at least one set of word-terms and (ii) at least one set of word-classes, wherein a term is one of a word-term and a word-class, where the plurality of terms selected is from the combination of terms, where those terms are in the combination of terms, where the terms assigned information-gain values are the combination of terms, where the plurality of terms selected is from the combination of

terms, where those terms are in the combination of terms, and where the terms are from the combination of terms, and wherein the combining results in a combination of terms

Diab teaches/suggests combining: (i) at least one set of word-terms and (ii) at least one set of word-classes, wherein a term is one of a word-term and a word-class ("word sense tagging... automatically sense annotating... large amounts of data... using an unsupervised algorithm... bootstrap... creating a sense-tagged corpus", Introduction, especially paragraph 3; "project the sense tags from the target side to the source side... KIND-OF-DRAMA sense... CALAMITY... the tagging... would yield... large number of French words will receive tags from the English sense inventory", Approach, especially 4th bullet, paragraph ending at the upper right of page 257, and last paragraph; Applicant does not claim that the word that the generation is based on was derived from the communication, so as long as the generated data includes the word that also exists in the communication, it is "based on the word" that the communication "comprises"; Diab teaches sense-tagging words to create a corpus of sense-tagged data. Each of these sense-tagged words includes a word-class like CALAMITY and a word-term like catastrophe. Diab teaches the existence of different senses [like KIND-OF-DRAMA] and it is at least obvious that catastrophe is not the only word with different senses in French. The sense-annotated words in the corpus, collectively, are a "combination of terms" because the classes/senses are combined with their corresponding words, and at the very least in set theory sets can include 1 element, or alternatively, collectively all of the sense-annotated words are a set of word-terms and their corresponding senses collectively are a set of word-classes)

where the plurality of terms selected is from the combination of terms, where those terms are in the combination of terms, where the terms assigned information-gain values are the combination of terms, where the plurality of terms selected is from the combination of terms, where those terms are in the combination of terms, and where the terms are from the combination of terms, and wherein the combining results in a combination of terms ("word sense tagging... automatically sense annotating... large amounts of data... using an unsupervised algorithm... bootstrap... creating a sense-tagged corpus", Introduction, especially paragraph 3; "project the sense tags from the target side to the source side... KIND-OF-DRAMA sense... CALAMITY... the tagging... would yield... large number of French words will receive tags from the English sense inventory", Approach, especially 4th bullet, paragraph ending at the upper right of page 257, and last paragraph; where Li [in Section 3, especially paragraph 2] teaches that the term-document matrix is generated from a labeled corpus, though it does not specifically state how the corpus is generated. Diab teaches generating a labeled corpus by a processor which contains the very information that Li wishes to extract for Li's matrix. Therefore, one of ordinary skill in the art can simply substitute the corpus used in Li with one generated from a processor as per Diab that contains the information needed to generate the matrix, and which can be used to bootstrap the classifier described in Li)

Therefore, it would have been obvious to one of ordinary skill in the art at the time of invention to perform a **simple substitution** of one corpus which the information-gain-matrix performs selection from with another, because Li teaches a classification

method/device **which differed from the claimed device by the substitution of a** corpus containing word class information and word term information generated by some means used by Li to generated a matrix, with another corpus containing the same information derived/generated by a processor. Diab teaches that a corpus generated by a processor and containing word class information and word term information **was known in the art. One of ordinary skill in the art could have substituted one** corpus **for another to obtain the predictable results** of a system which performs classification using a matrix generated from a corpus containing terms and classes (as per Li) where the corpus containing terms and classes is generated by a processor (as per Diab).

Li, in view of Diab, fail to teach wherein an automatic word class clustering algorithm is utilized to generate the word-classes.

Sakai teaches wherein an automatic word class clustering algorithm is utilized to generate the word-classes ("category decision rules... each text is classified to a category according to the category decision rule", col. 3, lines 35-50; "automatically creates a new category", col. 6, line 53 – col. 7, line 5; "if a cluster consisting of a large number of texts... new category to which this cluster is classified", col. 6, lines 34-40; "cluster generation unit", col. 6, lines 7-24; where the clustering is automatically performed and whose results is used for a new word class, and so it is an automatic word class clustering algorithm and is used to generate new word class [i.e., category] rules/information. Li and Diab teach where categories/senses are generated somehow [since they must have been derived from somewhere to be used in tagging], without

providing the specifics. Sakai teaches another method for generating the same data and so a simple substitution of the generation can be performed to yield the class/sense information used in Diab's sense tagging).

Therefore, it would have been obvious to one of ordinary skill in the art to perform a **simple substitution** of one category with another, because Li and Diab teach a **device which differed from the claimed device by the substitution of** categories generated by some unspecified manner with categories generated by clustering. Sakai teaches that categories generated by clustering **were known in the art. One of ordinary skill in the art could have substituted one known element for another** by using categories generated from clustering instead in the unsupervised tagging in Diab **in order to obtain the predictable results of** a system that performs classification based on a matrix derived from word class and word term data (Li) where the word class and word term data are automatically generated by a processor (Diab) based on classes determined initially using a form of clustering (Sakai).

As per Claim 13, Li teaches wherein the selected terms in the plurality of terms are processed by the processor-based device to form a term-category matrix from which a joint classifier determines at least one category for the word, and wherein the processor-based device comprises the joint classifier ("LSI classifier... categorize an unknown document... derived from... as in LSI according to IG enhanced term-document matrix... similarity... n-best categories", Section 3; "user's request", Section

4; where categorizing a document by consequence categorizes the word in that document, and this is "joint classification" in the sense that it "jointly" uses both term information and category information to perform classification [i.e. "joint classifier is configured to determine at least one category for the words, by applying a combination of word information and word class information to the words", Specification, page 6]).

As per Claim 14, Li teaches generating by the processor-based device a term-category matrix, wherein (i) the term-category matrix comprises a plurality of terms and a plurality of categories, and (ii) each term in the matrix is associated with at least one category ("term-document matrix M is formed by terms... selected term is mapped to a unique row vector and each category is mapped to a unique column vector", Section 3, especially paragraph 3; where the matrix is formed by combining terms [i.e. word terms] and categories [word-classes], and in a matrix, the matrix cell corresponding to a specific row's term/word-term and a specific column's category/word-class associates the term and the category corresponding to a cell; Alternatively, words are naturally associated with some particular class [e.g. words that are verbs, medical words, English words, etc.] and "at least one category" as claimed does not necessarily refer to any category in particular so any word naturally reads on this claim limitation because words are, by virtue of what they are, part of some form of category)

determining from the term-category matrix, based on a joint classification of the word by the processor-based device, the category for the word ("LSI classifier... categorize an unknown document... derived from... as in LSI according to IG enhanced

term-document matrix... similarity... n-best categories", Section 3; "user's request", Section 4; where categorizing a document by consequence categorizes the word in that document, and this is "joint classification" in the sense that it "jointly" uses both term information and category information to perform classification [i.e. "joint classifier is configured to determine at least one category for the words, by applying a combination of word information and word class information to the words", Specification, page 6])

routing the communication by a communication system to a particular one of a plurality of destination terminals of the communication system, wherein the routing is based on the category of the word, and wherein the communication system comprises the processor-based device and the plurality of destination terminals ("routing... appropriate destination within a call center", Section 4; where the routing to destinations in the call-center/system based on the query's categorization, which is done via the joint classifier, which includes categorizing the words in the query, and the "system" can be interpreted as the call router and all of the places that the call is routed to).

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to ERIC YEN whose telephone number is (571)272-4249. The examiner can normally be reached on M-F 7:30-4:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on 571-272-7602. The fax phone

number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

EY 1/28/11
/Eric Yen/
Primary Examiner, Art Unit 2626